

AI-Assisted Identification and Quantification of Pharmaceutical Impurities using HPLC Chromatographic Data

J. Abdul Basith^{1*}, V. Brindha Shree¹, J. Balagopal¹, P. Dharshini¹, I. Jai Surya¹, Ravisankar Mathesan¹, Nataraj Palaniyappan²

^{1*,1}Srinivasan College of Pharmaceutical Sciences, Trichy

²Scientist, Novitium Pharma LLC, New jersey, USA

*Corresponding E-mail: jafarbasith8@gmail.com

Keywords

Artificial Intelligence (AI),
High-Performance Liquid
Chromatography (HPLC),
Pharmaceutical Impurities,
Chemometrics,
Machine Learning (ML)

Received 09-02-2026

Revised 26-02-2026

Accepted 21-03-2026

Published 25-05-2026

ABSTRACT

Pharmaceutical impurities present in active pharmaceutical ingredients (APIs) and finished drug products can have profound impacts on drug safety, efficacy, and stability. Regulatory authorities such as the International Council for Harmonisation (ICH), U.S. Food and Drug Administration (FDA), and European Medicines Agency (EMA) require stringent impurity identification and quantification to ensure drug quality and patient safety. High-Performance Liquid Chromatography (HPLC) remains the most common analytical method for impurity profiling, thanks to its high sensitivity, precision, and ability to separate complex mixtures. However, traditional HPLC data interpretation relies heavily on manual analysis, making it time-consuming, operator-dependent, and prone to human error. Artificial Intelligence (AI) and Machine Learning (ML) have emerged as transformative tools for automating chromatographic data analysis, enabling rapid and accurate impurity detection, identification, and quantification. AI algorithms can efficiently process large chromatographic datasets, identify subtle peak patterns, correct baseline drifts, and predict retention behaviours under various chromatographic conditions. This comprehensive article explores the integration of AI approaches into HPLC impurity profiling, covering theoretical foundations, data preprocessing, peak recognition, automated quantification, and future implications for pharmaceutical quality control and regulatory compliance.

INTRODUCTION

The pharmaceutical industry upholds stringent quality standards to ensure that every drug product released to the market is safe, effective, and stable. A critical part of this mandate involves the control and characterization of impurities unwanted chemical entities that arise from synthetic reactions, intermediate by-products, degradation pathways, or contamination during formulation and storage. Even trace-level impurities can jeopardize the therapeutic performance or induce toxicity, underscoring their importance in quality control and regulatory compliance. Regulatory frameworks such as ICH Q3A (for drug substances), Q3B (for drug products), Q3C (for residual solvents), Q3D (for elemental impurities), and M7 (for mutagenic impurities) define specific guidelines for impurity profiling, establishing thresholds for detection, identification, and qualification.[1-9] The thresholds are particularly strict for genotoxic and carcinogenic impurities, which must often be controlled at sub-parts-per-billion levels. High-Performance Liquid Chromatography (HPLC) serves as the cornerstone of impurity analysis. It offers high resolution, reproducibility, and adaptability for analysing complex pharmaceutical matrices.[10] However, as APIs and formulations increase in structural complexity, traditional chromatographic interpretation performed manually through visual inspection of retention times

and peak areas has become inefficient and inconsistent. Artificial Intelligence (AI) introduces the possibility of intelligent, automated chromatographic data analysis. Utilizing machine learning (ML), neural networks (ANNs), and deep learning (DL), AI systems can process large datasets, identify complex patterns in chromatographic outputs, and make data-driven predictions for impurity identification and quantification. These methods reduce subjectivity, enhance reproducibility, and shorten analytical timelines across all stages of drug development.[11]

Pharmaceutical Impurities and Their Importance

Pharmaceutical impurities are any undesired chemical substances present in a drug substance, excipient, or drug product that are not active ingredients but result from synthesis, degradation, or contamination.

Classification of Impurities:

1. Organic Impurities:

Derived from synthesis and degradation processes, they include starting materials, process intermediates, by-products, and degradation products.

2. Inorganic Impurities:

Arise from reagents, catalysts, heavy metals; filter aids, or other inorganic substances used in the manufacturing process.

3. **Residual Solvents:** Volatile organic chemicals employed during synthesis or crystallization processes that remain in trace quantities in the final product.

Each type poses different risks. For instance, organic impurities may affect pharmacological activity or toxicity, inorganic impurities pose heavy-metal toxicity risks, and residual solvents can impact drug stability and organ health. Regulatory bodies establish impurity thresholds based on toxicological evaluation for example, impurities above **0.05%** must generally be reported, and those above **0.10%** identified for most APIs. For highly potent or genotoxic impurities, limits are far lower, typically within nanogram-per-gram ranges.[1-3]

Artificial Intelligence in Chromatographic Data Analysis

AI is revolutionizing analytical chemistry by enabling the interpretation of complex datasets beyond human cognitive limits. In HPLC, AI algorithms can manage noisy datasets, extract subtle peak features, and handle nonlinear relationships between chromatographic variables.

Common AI and ML Techniques Used in HPLC

1. **Artificial Neural Networks (ANNs):** ANNs simulate the neural structure of the brain, enabling models to learn non-linear relationships between chromatographic parameters like solvent composition, flow rate, pH, and column temperature—and observed retention or resolution outcomes. They are particularly effective in predicting retention times and optimizing separation conditions.[12]
2. **Support Vector Machines (SVMs):** SVMs are used for classifying and recognizing patterns in chromatographic data. They can identify unknown compounds based on peak features, facilitating impurity identification even when standards are unavailable.
3. **Random Forests and Decision Trees:** These ensemble models analyze multiple input variables such as retention time, peak area, and intensity to categorize chromatographic peaks. They are especially helpful for automated detection of outlier peaks potential impurities.
4. **Deep Learning (CNNs and RNNs):** Deep learning networks can extract hierarchical features directly from raw chromatographic signals, eliminating the need for manual feature extraction. Convolutional Neural Networks (CNNs) can detect peak edges and shapes, while Recurrent Neural Networks (RNNs) handle time-series chromatographic data, improving analysis of complex or co-eluted peaks.[13]

Integrated properly, these algorithms can decrease human intervention, enhance resolution optimization, and improve reproducibility of chromatographic workflows.

AI-Assisted Identification of Pharmaceutical Impurities

AI identifies impurities by analyzing the shape, intensity, and retention time of chromatographic peaks, often leveraging spectral data from diode array detectors (DAD) or coupled mass spectrometry.

Key AI-assisted identification tasks include:

- **Peak Recognition and Classification:** AI algorithms perform automated baseline correction, noise

reduction, and detect abnormal peaks characteristic of impurities.

- **Spectral Deconvolution:**

Through chemometric approaches like Independent Component Analysis (ICA) or Multivariate Curve Resolution (MCR), AI separates co-eluted peaks into their individual components.

- **Pattern Recognition and Retention Prediction:**

Using structure–property relationships, AI models can correlate molecular descriptors to retention behavior under differing chromatographic conditions.

This predictive ability helps analysts anticipate impurity elution, optimize column selection, and avoid co-elution with main analytes.

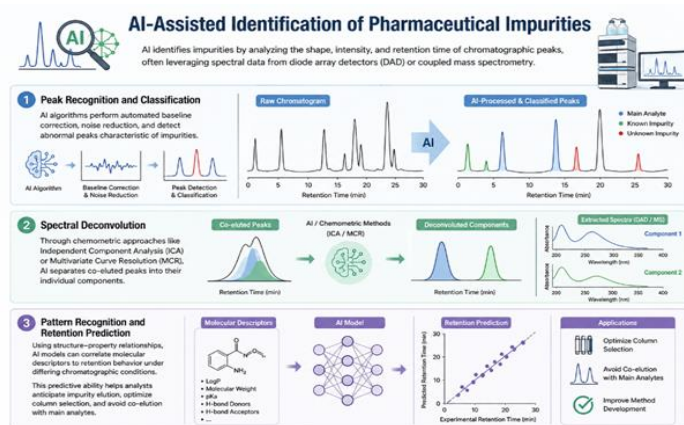


Figure 1: AI Assisted Identification of Pharmaceutical Impurities

Quantification of Impurities Using AI-Assisted HPLC Analysis: Traditional impurity quantification depends on calibration curves and manual integration of chromatographic peaks. However, these methods can be inaccurate for trace impurities or those without reference standards. AI methods improve accuracy through automation and intelligent modeling.[14]

AI Enhancements in Quantification

- Automated and consistent peak detection/integration
- Baseline drift correction and noise filtering via ML algorithms
- Retention time prediction for unknown impurities
- Use of chemometric modeling for impurity quantification without calibration standards

Chemometric and Statistical Approaches

Principal Component Analysis (PCA): PCA deconstructs spectral data into uncorrelated components, revealing patterns and permitting impurity quantification through summed score chromatograms, which better represent true concentrations than single-wavelength measurements.[15]

Multivariate Curve Resolution–Alternating Least Squares (MCR-ALS):

This chemometric method separates overlapping peaks into pure component profiles and spectra, enabling quantification of impurities even in overlapping chromatographic zones.

Machine Learning for Response Factor Estimation:

When some impurities have reference standards, regression or neural models can learn structure–response relationships,

estimating relative response factors for new impurities. This reduces dependency on laborious standard synthesis. Research indicates that AI integration in chromatographic quantification can cut analysis time by up to 40% while maintaining high accuracy crucial for rapid regulatory decision-making and manufacturing control.[16]

Applications in Pharmaceutical Quality Control

AI-assisted HPLC analysis offers value across all stages of pharmaceutical development and manufacturing:

- 1. Drug Development and Stability Studies:**
AI tools help predict impurity formation during forced degradation or stress testing, expediting stability-indicating method validation.
- 2. Manufacturing Process Monitoring:**
Integration with real-time process analytical technology (PAT) allows ML systems to detect impurities during production, enabling immediate corrective actions.
- 3. Quality Control Laboratories:**
AI automates peak interpretation, ensuring consistent and repeatable impurity quantification while reducing operator workload.
- 4. Regulatory Documentation:**
AI-driven impurity analysis enhances traceability, providing structured digital evidence of impurity control for compliance with ICH and FDA guidelines.

Regulatory Considerations and Quality Assurance

The regulatory acceptance of AI-assisted impurity profiling hinges on rigorous validation and transparency. AI-enhanced methods must conform to traditional analytical validation parameters specificity, accuracy, precision, linearity, limits of detection and quantification, and robustness.

Key regulatory considerations include:

- **Model Transparency:**
- Regulatory agencies expect clear documentation of algorithm logic, training datasets, and performance metrics.
- **Data Integrity and Compliance:** AI systems must comply with 21 CFR Part 11 and Annex 11 standards, ensuring traceable and secure data management.⁹
- **Lifecycle Management:** Under ICH Q14 and Q2(R2), AI models used in analytical procedures must follow a lifecycle approach, with controlled retraining, revalidation, and change documentation.¹⁻⁸

Regulators acknowledge the efficiency of AI in analytical QbD frameworks, if systems remain auditable and scientifically justified.

Case Studies and Practical Implementations

1. Comprehensive Orthogonal Method Evaluation Technology (COMET)

COMET is an AI-assisted platform that screens drug samples using multiple orthogonal HPLC methods with diverse selectivity's. By leveraging chemometric algorithms and mass spectral data, it automatically identifies corresponding peaks across all chromatograms, reducing method evaluation time from weeks to hours.[18]

2. PCA-Based HPLC-DAD Quantification

Principal Component Analysis applied to diode-array chromatograms has demonstrated superior accuracy in impurity quantification compared to conventional single-wavelength integration especially for co-eluting or spectrally similar compounds.[19]

3. Predictive Modeling for Genotoxic Impurity Control

Quantitative Structure–Activity Relationship (QSAR) and ML models can predict the genotoxic potential of impurities before experimental detection, streamlining analytical prioritization and reducing time to compliance.[20]

Challenges and Limitations [21]

Despite its potential, AI-assisted impurity analysis faces practical limitations:

- **Data Quality and Quantity:** Insufficient or biased training datasets reduce prediction accuracy.
- **Model Interpretability:** Complex deep learning models are often opaque (“black box”), complicating validation.
- **Integration Barriers:** Implementing AI solutions requires infrastructure investments and analyst training.
- **Detection Limits:** Ultra-trace impurities still pose analytical challenges, requiring instrument-based improvements.
- **Unknown Structural Elucidation:** AI helps detect but not identify chemical structures; LC-MS/MS or NMR remains necessary for confirmation.

Future Directions

AI's integration into pharmaceutical impurity analysis is poised to evolve further toward continuous, intelligent quality control systems:

Real-Time PAT Integration: AI-augmented chromatographic monitoring for continuous manufacturing environments.

Automated Spectral Interpretation: Combining AI with LC-MS, IR, and NMR data to automatically propose impurity structures.

Predictive Degradation Modeling: Anticipating impurity formation kinetics and degradation pathways for formulation design.

Green Chemistry Optimization: Minimizing solvent consumption and analysis time via AI-driven modeling.

Federated Learning:

Enabling multi-organization collaboration while maintaining data confidentiality, improving model robustness.[22-25]

Table.1: Software Platforms Supporting AI in Chromatographic Analysis [17]

Software	AI/Analytics Role	Primary Application	Benefits for Method Development
MATLAB®	ML, ANN, DL modeling	Statistical and chemometric modeling	Captures non-linear chromatographic relationships
Python (scikit-learn, TensorFlow, Keras)	ML/DL frameworks	Custom model development and visualization	Deep learning-based prediction, automation pipelines
R Programming	Statistical and ML analysis	Data preprocessing, quantitative modeling	Enhanced reproducibility and exploratory data analysis
Design-Expert®	ML-assisted experimental design	Optimization of analytical parameters	Reduces trial count, incorporates QbD
Empower®, Chromeleon®, OpenLab®	Integration with AI modules	Data acquisition and chromatographic control	Automated peak detection, improved integration accuracy

These software systems underpin AI-enhanced HPLC method development, allowing direct data integration from chromatographic instruments into ML models for continuous learning and optimization.

CONCLUSION

AI-assisted chromatographic analysis represents a transformative leap in pharmaceutical impurity profiling. By integrating machine learning, deep learning, and chemometric modeling with HPLC data interpretation, analysts can achieve unprecedented precision, automation, and efficiency. These AI-driven innovations not only streamline impurity detection and quantification but also significantly reduce analysis time, enhance reproducibility, and minimize human error and subjective bias. Beyond routine analysis, the true potential of AI lies in its ability to enable predictive and proactive quality control. Advanced algorithms can learn from historical chromatographic data, identify hidden patterns, and forecast impurity trends under varying formulation, storage, or manufacturing conditions. This shift from reactive to predictive analytics empowers pharmaceutical scientists to address potential quality issues before they arise, thereby strengthening risk management strategies and ensuring regulatory compliance. Moreover, the integration of AI with modern analytical instrumentation supports real-time monitoring and continuous manufacturing, aligning with emerging concepts such as Quality by Design (QbD) and Process Analytical Technology (PAT). As global regulatory authorities gradually recognize and incorporate AI-driven methodologies into their guidelines, the adoption of such technologies will become increasingly widespread and indispensable. In the long term, AI-assisted chromatographic systems will not only enhance analytical performance but also contribute to cost efficiency, sustainability, and faster drug development timelines. Ultimately, this convergence of artificial intelligence and analytical science ensures that pharmaceutical products maintain the highest standards of safety, purity, and efficacy throughout their lifecycle, marking a significant advancement in modern pharmaceutical quality assurance.

CONFLICT OF INTERESTS

The authors declare no conflict of interest

ETHICS APPROVAL: Not applicable

FUNDING: This study received no specific funding from public, commercial, or not for profit funding agencies.

AI TOOL DECLARATION

The authors declare that no AI and related tools are used to write the scientific content of this manuscript.

DATA AVAILABILITY

Data will be available on request

REFERENCES

- [1] International Council for Harmonisation. ICH guideline Q3A(R2): impurities in new drug substances. Geneva: ICH; 2006.
- [2] International Council for Harmonisation. ICH guideline Q3B(R2): impurities in new drug products. Geneva: ICH; 2006.
- [3] International Council for Harmonisation. ICH guideline Q3C(R8): residual solvents. Geneva: ICH; 2021.
- [4] International Council for Harmonisation. ICH guideline Q3D(R2): elemental impurities. Geneva: ICH; 2022.
- [5] International Council for Harmonisation. ICH guideline M7(R1): assessment and control of DNA reactive (mutagenic) impurities in pharmaceuticals. Geneva: ICH; 2017.
- [6] International Council for Harmonisation. ICH guideline Q2(R2): validation of analytical procedures. Geneva: ICH; 2023.
- [7] International Council for Harmonisation. ICH guideline Q14: analytical procedure development. Geneva: ICH; 2023.
- [8] U.S. Food and Drug Administration. Guidance for industry: analytical procedures and methods validation for drugs and biologics. Silver Spring (MD): FDA; 2015.
- [9] European Medicines Agency. Guideline on the limits of genotoxic impurities. London: EMA; 2014.
- [10] LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature*. 2015;521(7553):436-44.
- [11] Daszykowski M, Walczak B. Use of artificial neural networks in chromatography. *J Chromatogr A*. 2006; 1109(1): 1-10.
- [12] Goodacre R, Vaidyanathan S, Dunn WB, Harrigan GG, Kell DB. Metabolomics by numbers: acquiring and understanding global metabolite data. *Trends Anal Chem*. 2004;23(4):245-55.
- [13] Wold S, Esbensen K, Geladi P. Principal component analysis. *Chemom Intell Lab Syst*. 1987;2(1-3):37-52.
- [14] Tauler R. Multivariate curve resolution applied to second-order data. *Chemom Intell Lab Syst*. 1995;30(1):133-46.
- [15] Bro R, Smilde AK. Principal component analysis. *Anal Methods*. 2014;6(9):2812-31.
- [16] Gromski PS, Muhamadali H, Ellis DI, Xu Y, Correa E, Turner ML, et al. A tutorial review: metabolomics and chemometrics in analytical chemistry. *Anal Chim Acta*. 2015;879:10-23.
- [17] Kovatcheva P, et al. Application of chemometrics in HPLC method development and validation. *J Chromatogr Sci*. 2012;50(5):403-10.
- [18] Singh YR, Shah DB, Kulkarni M, Patel SR, Maheshwari DG, Shah JS, et al. Current trends in chromatographic prediction using artificial intelligence and machine learning. *Anal Methods*. 2023;15:2785-97.
- [19] Nikfarjam A, Sarker A, O'Connor K, Ginn R, Gonzalez G. Pharmacovigilance from social media: mining adverse drug reaction mentions using sequence labeling. *J Am Med Inform Assoc*. 2015;22(3):671-81.
- [20] Zhang Z, et al. Machine learning in chromatographic analysis: a review. *J Pharm Biomed Anal*. 2020; 186: 113251.

- [21] Tantawy MA, Wahba IA, Saad SS, et al. Classical versus chemometric tools for spectrophotometric determination of drug impurities. *BMC Chem.* 2023; 17:49.
- [22] Alves E, Gurupadayya BM, Prabhakaran P. Artificial intelligence in HPLC method development: a critical review. *Crit Rev Anal Chem.* 2025.
- [23] Divekar K, Prasad SK. Automation in analytical chemistry: role of AI in chromatography. *Int J Appl Pharm.* 2024; 16(3):1-10.
- [24] Artificial intelligence for method development in liquid chromatography. *Trends Anal Chem.* 2025; 192: 118320.
- [25] Niessen WMA. Progress in liquid chromatography-mass spectrometry instrumentation and its impact on impurity profiling. *J Chromatogr A.* 2003;1000(1-2):413-36.
- [26] Snyder LR, Dolan JW. High-performance gradient elution: the practical application of the linear-solvent-strength model. *J Chromatogr A.* 2007;1157(1-2):3-25.